

DEEP ROB

Lecture 24

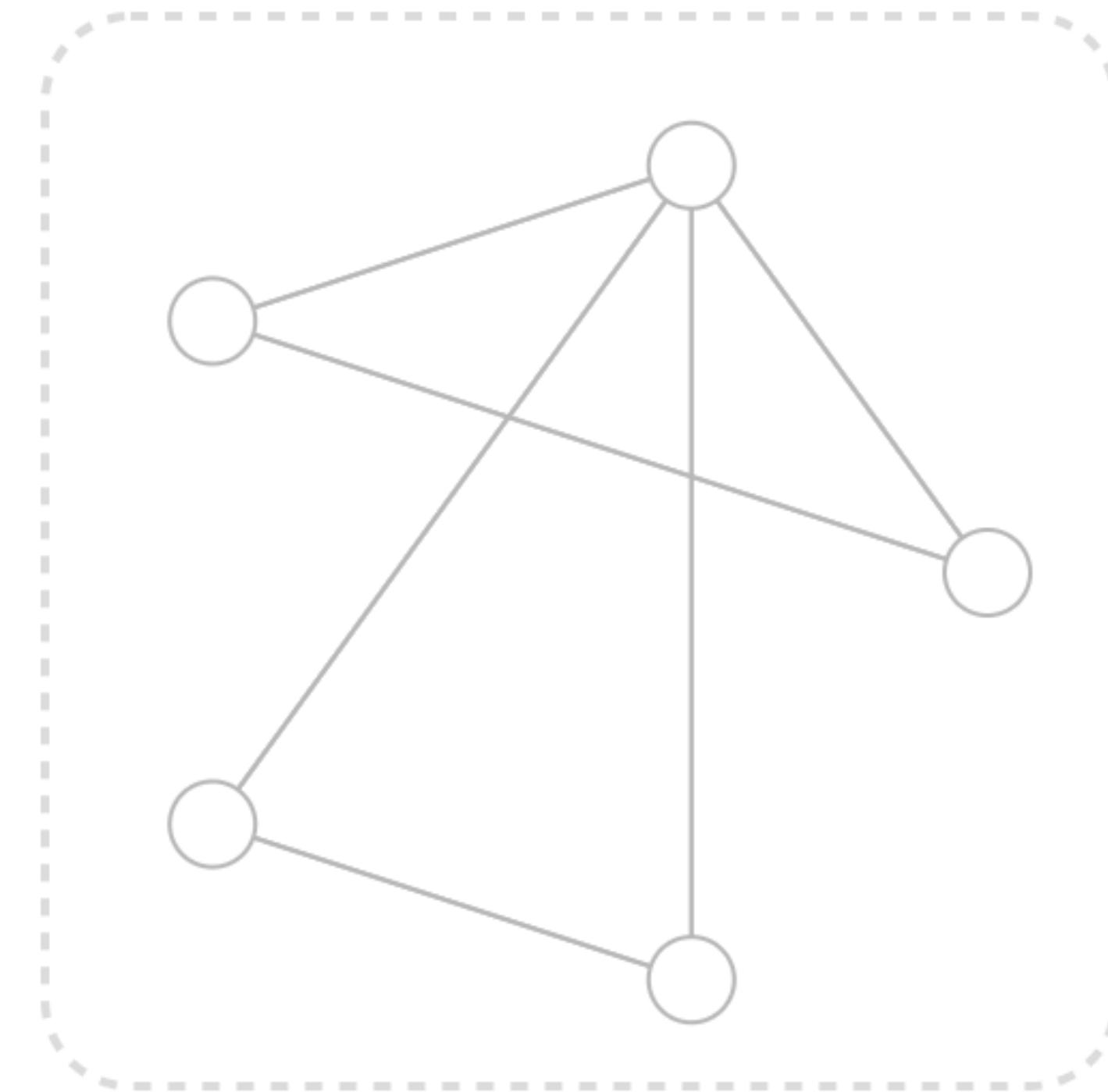
Semantic Scene Graphs

University of Michigan | Department of Robotics





Graphs

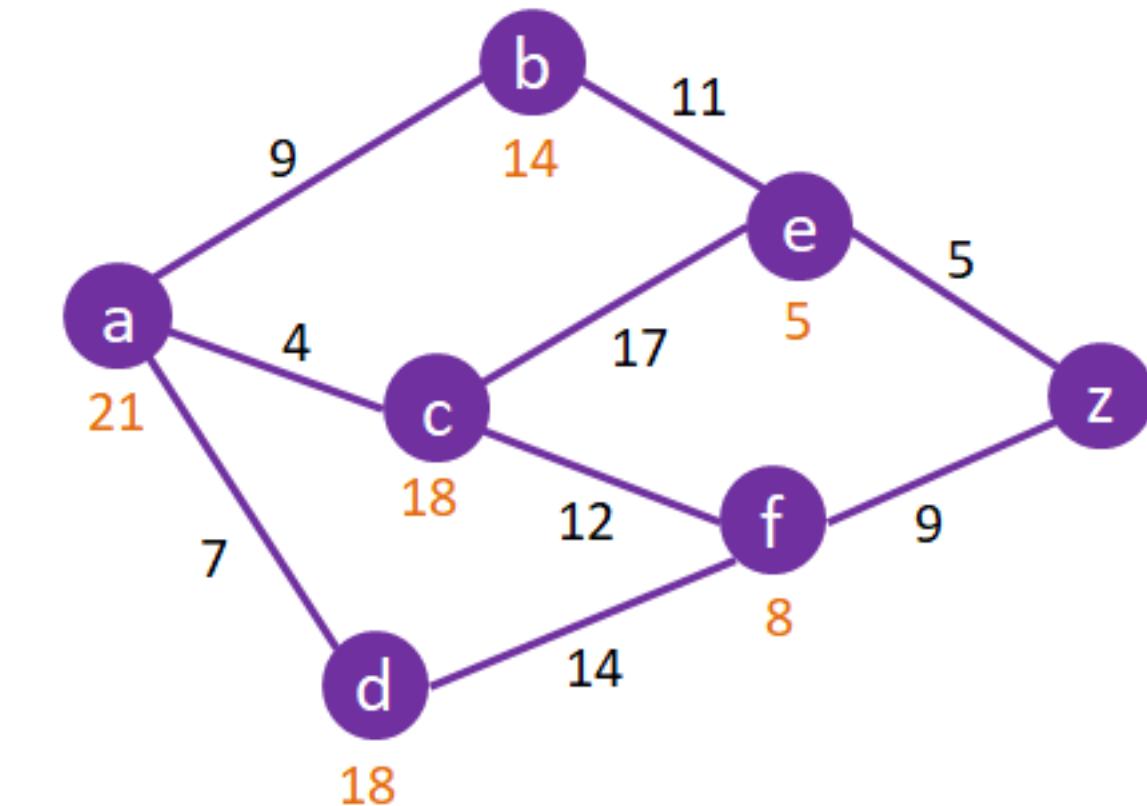


nodes

edges

Examples:

TSP
A*
GNNs, GCNs





“Computer vision was focused on **disconnected** objects” -- Ranjay Krishna

Image Classification



Object Detection



Instance Segmentation





How can we label images?

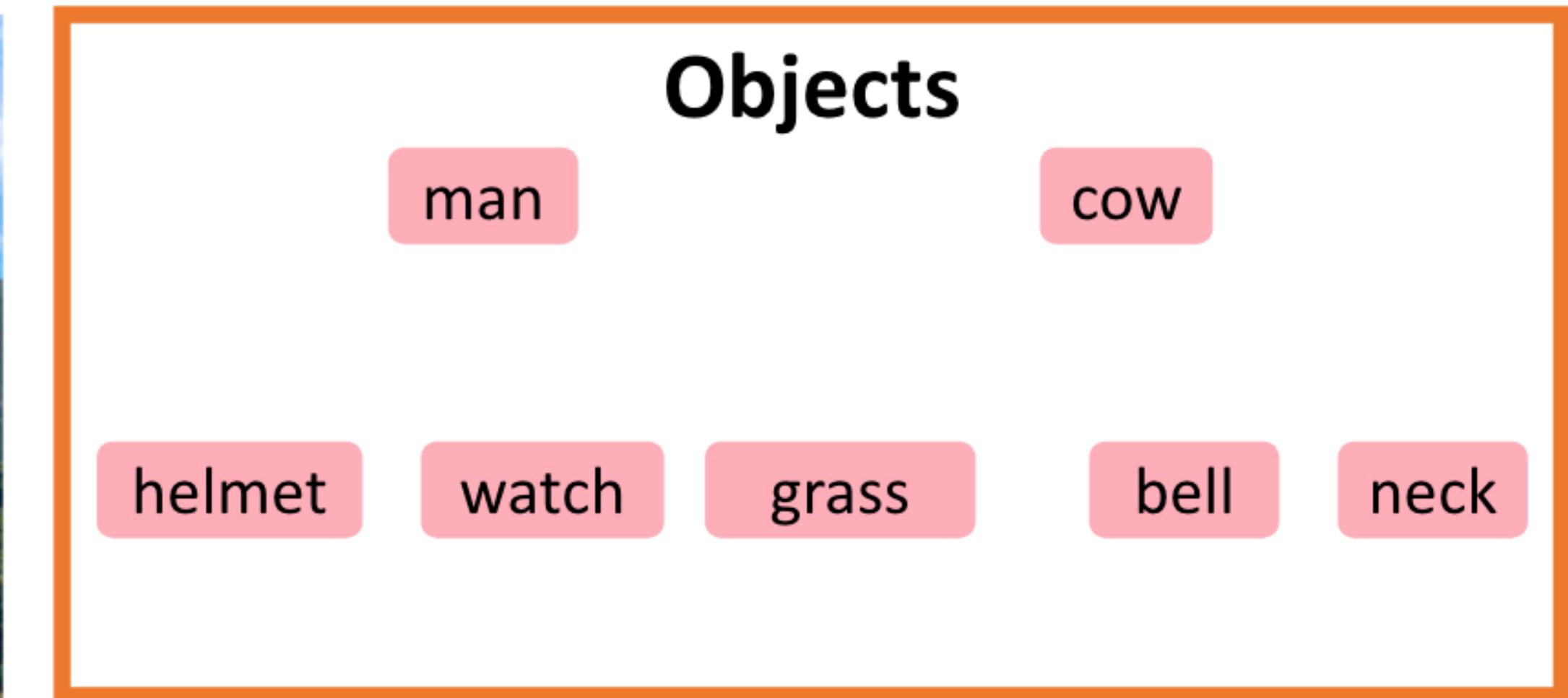


Image Label

Alp

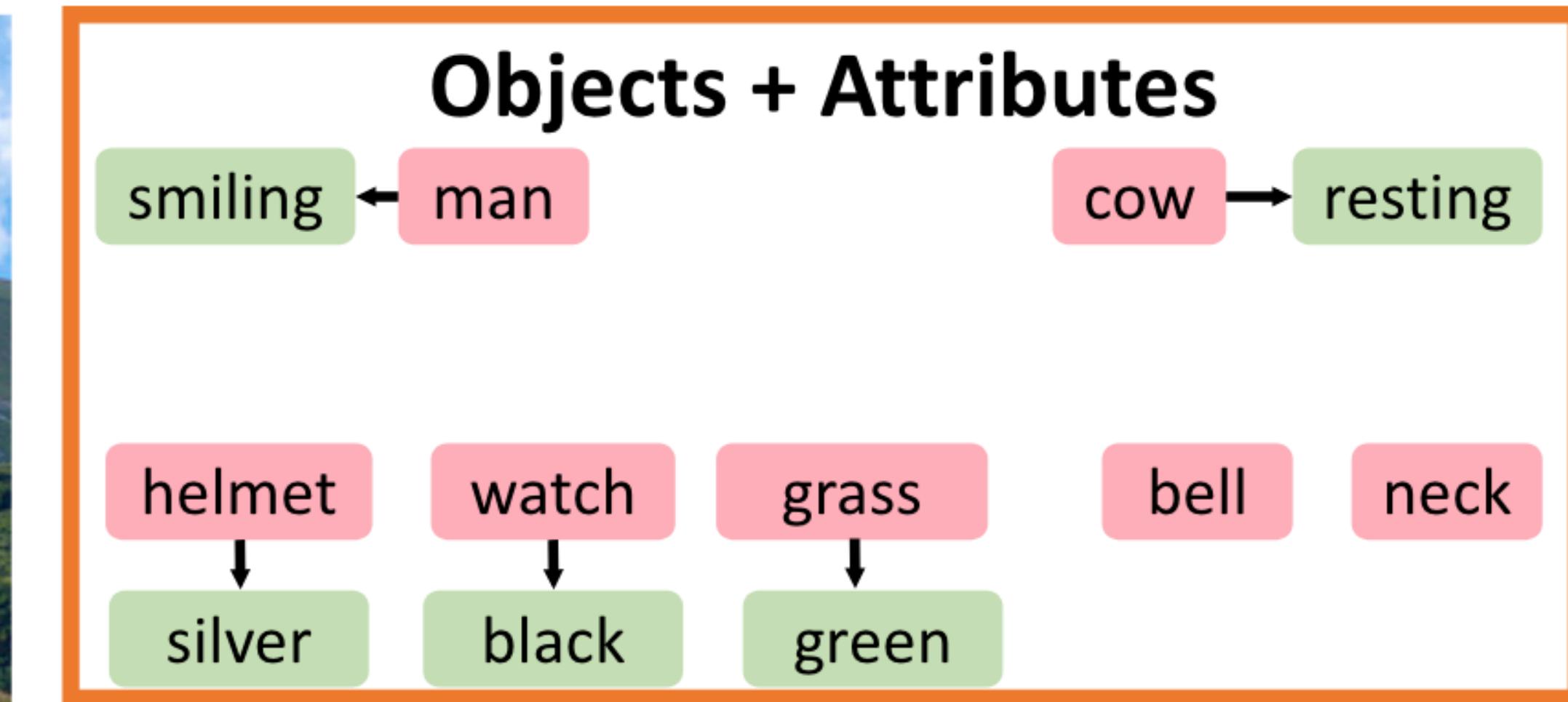


How can we label images?



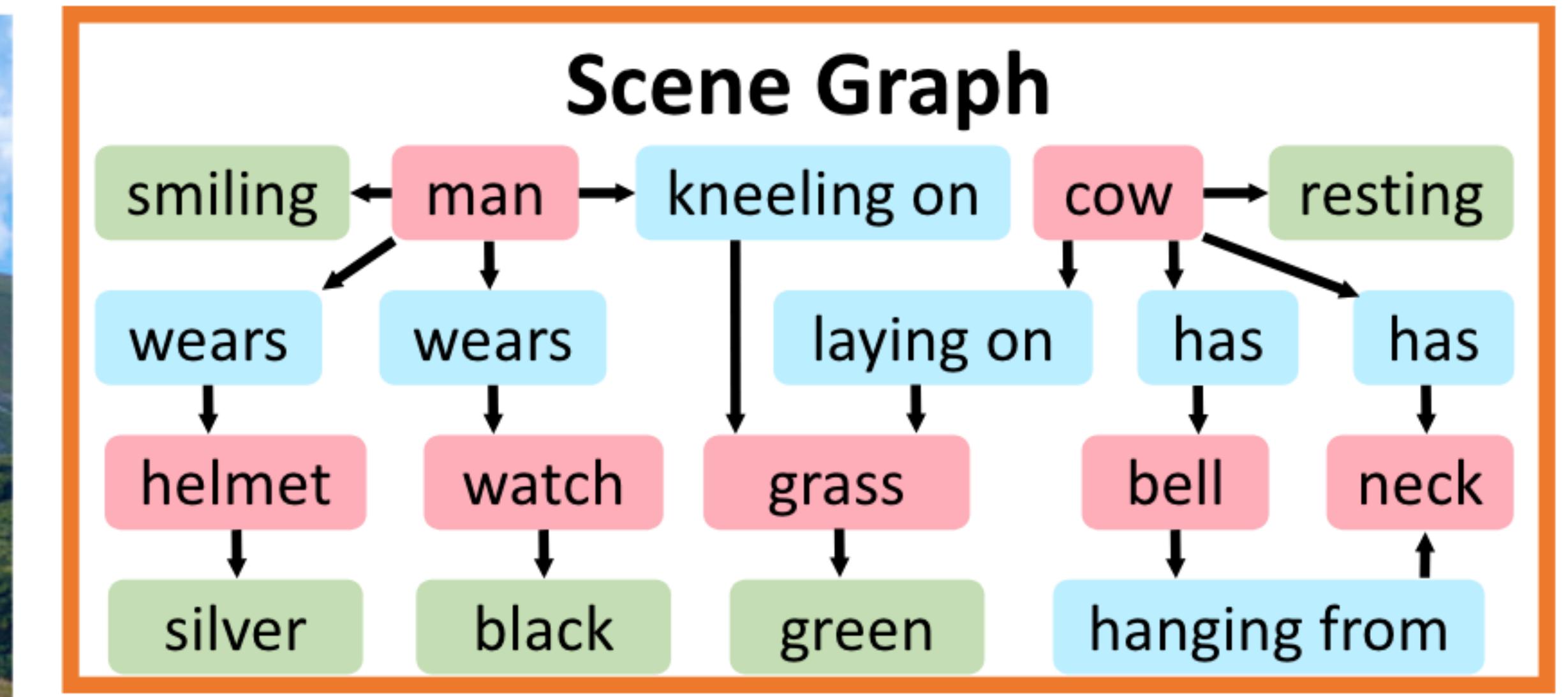


How can we label images?





How can we label images?

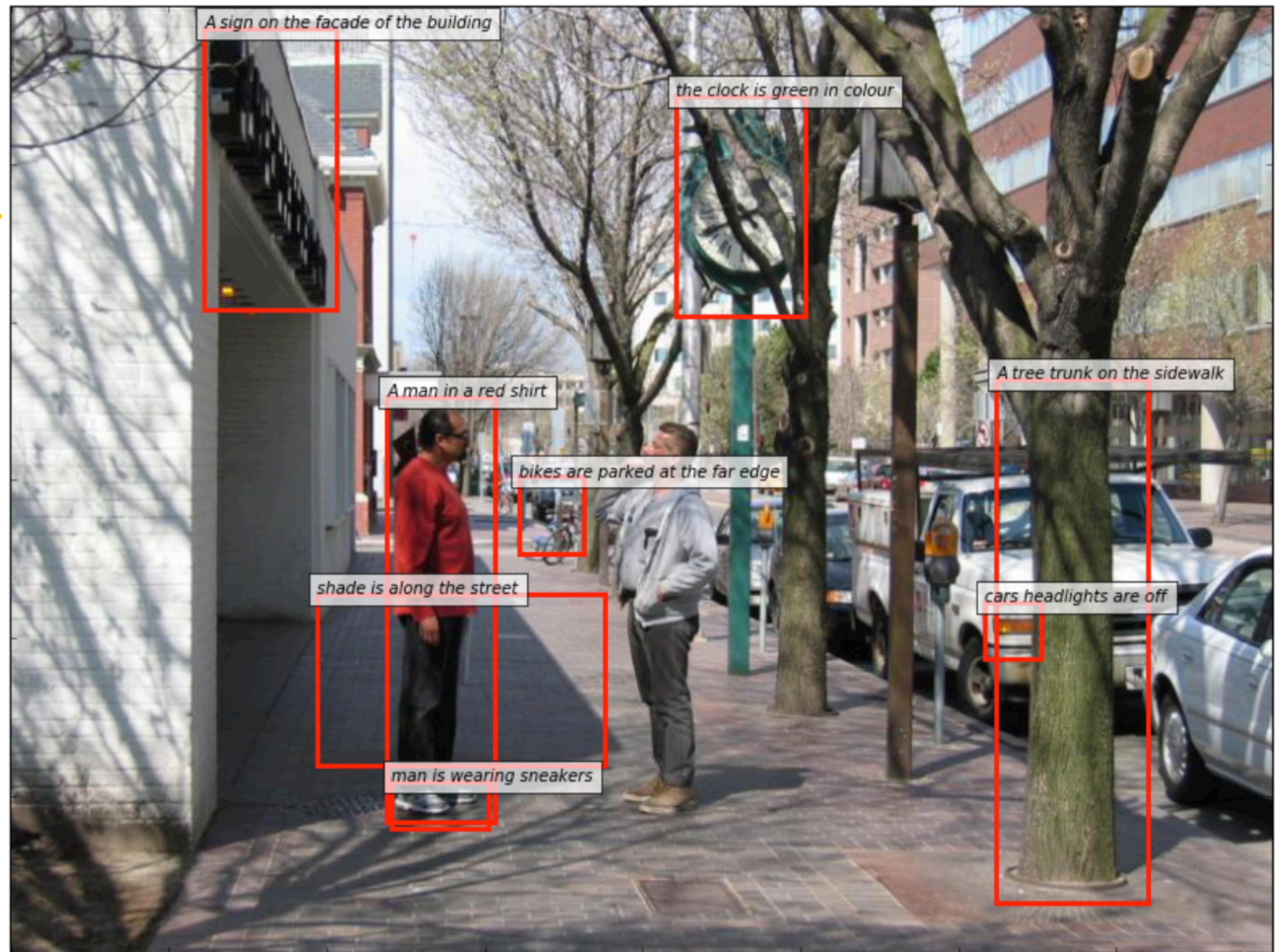


Describe **rich** visual knowledge



 VISUALGENOME

Visual Genome Dataset





Questions

1. Q: What is this person wearing on his head? A: A helmet.

2. Q: Where are clouds? A: In the sky.

3. Q: What color is the cow? A: Brown.

4. Q: What is green? A: Grass.

5. Q: What color is the sky? A: Blue.

6. Q: Who is wearing a helmet? A: A man.

7. Q: Where is the helmet? A: Man's head.

8. Q: Where are the hills? A: The distance.

9. Q: What is man doing? A: Kneeling.

10.Q: What is man wearing? A: Cycling gear.

11.Q: What color are clouds? A: White.

12.Q: Where are man's knees? A: In grass.

13.Q: Where is the flag? A: On shirt.

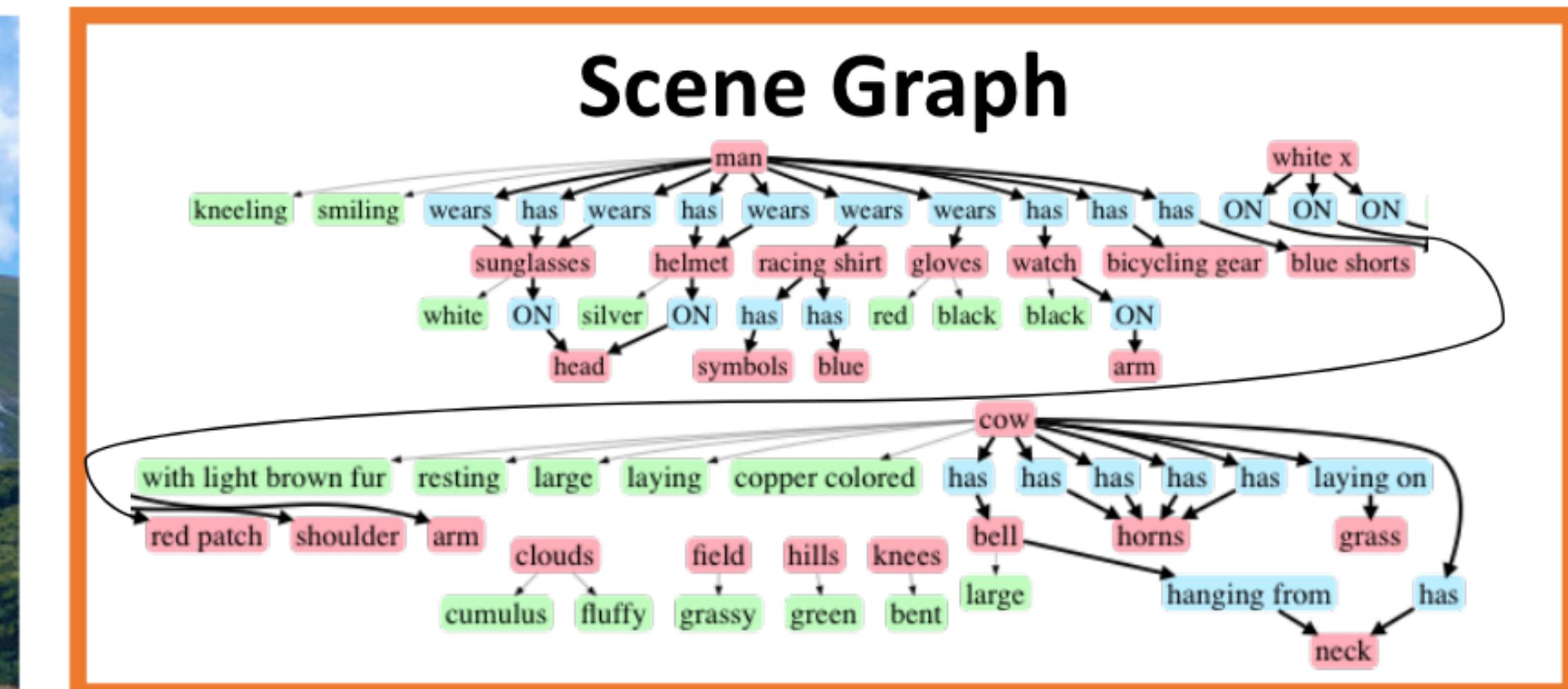
14.Q: What is the cow doing? Laying.

15.Q: What color is cow fur? A: Brown.

16.Q: What is red and white? A: Man's shirt.

17.Q: What is green? A: Grass.

18.Q: Who has horns? A: The animal.

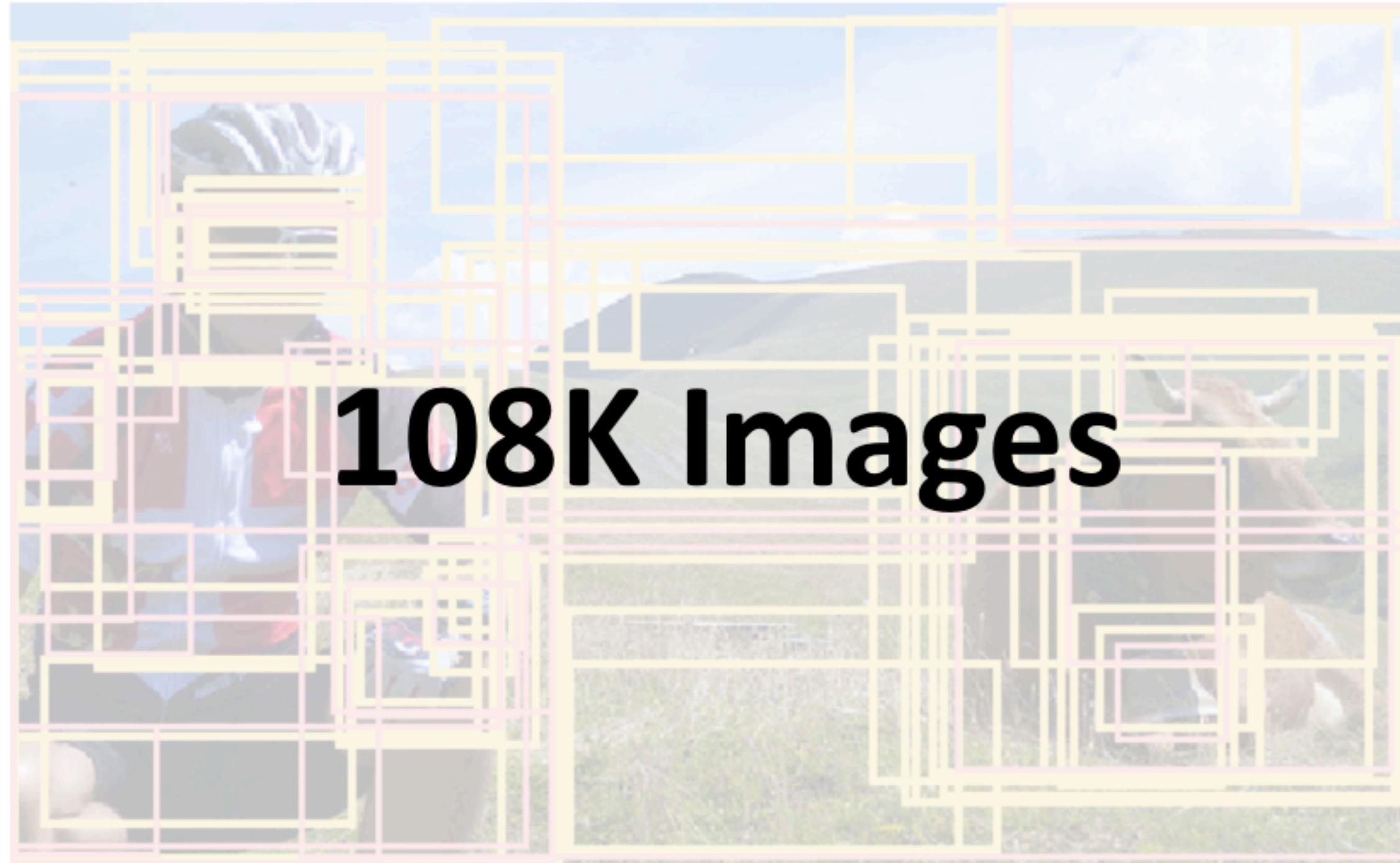


Region Descriptions

- | | | | | | |
|-----|---|-----|--|-----|-------------------------------------|
| 1. | blue ski with clouds | 17. | White gloves with red and black colors | 34. | man with knees bent on grass |
| 2. | man's sunglasses | 18. | Many hills in the distance | 35. | shirt on man with flag and symbols |
| 3. | roaming hills | 19. | sunglasses on a mans face | 36. | sport sunglasses on man's head |
| 4. | grassy field | 20. | A man kneeling down | 37. | face of man smiling |
| 5. | man's biking helmet | 21. | large brown cow | 38. | a cow laying down resting on grass |
| 6. | a gig cow bell | 22. | large metal bell on cow's collar | 39. | horns on head of cow |
| 7. | cow laying next to the man | 23. | short horn's on cow's head | 40. | cow with light brown fur |
| 8. | man wearing Denmark racing shirt | 24. | man wearing bicycling gear | 41. | there are mountains in background |
| 9. | man's whites biking gloves | 25. | silver bike helmet | 42. | clouds in the sky |
| 10. | man's watch | 26. | white rimmed sunglasses | 43. | man is smiling |
| 11. | Big fluffy clouds in the sky | 27. | red and white cross on man's shoulder | 44. | the shirt is red white and blue |
| 12. | A large bell hanging from a cows neck | 28. | rolling green hills | 45. | he has on blue shorts |
| 13. | A copper colored cow laying in the grass with horns | 29. | wrist watch on man's arm | 46. | animal has a collar around his neck |
| 14. | The horns from a cow | 30. | white cumulus clouds in a blue sky | 47. | the grass is green |
| 15. | A mans head who is wearing a helmet and glasses | 31. | a biking helmet on head of man | 48. | Man is wearing glasses |
| 16. | A white X on red patch | 32. | a glove on man's hand | 49. | there is a cross on the arm |
| | | 33. | a black watch on man's wrist | 50. | this animal has horns |



Visual Genome: A Vision + Language Dataset



108K Images

- # 1.7M Questions

1.7M Questions



5.4M Region Descriptions

5.4M Region Descriptions

Descriptions



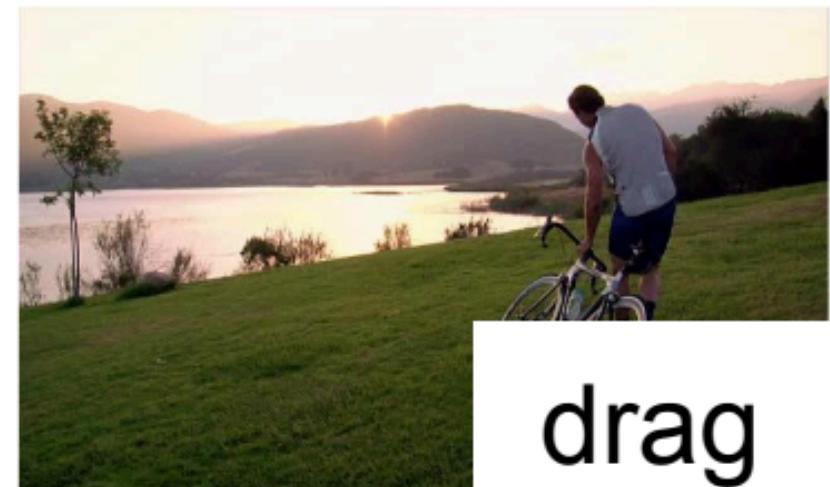
Challenges

- Challenge 1: Quadratic Explosion

- N objects,
- K relationships

leading to **N²K** detectors

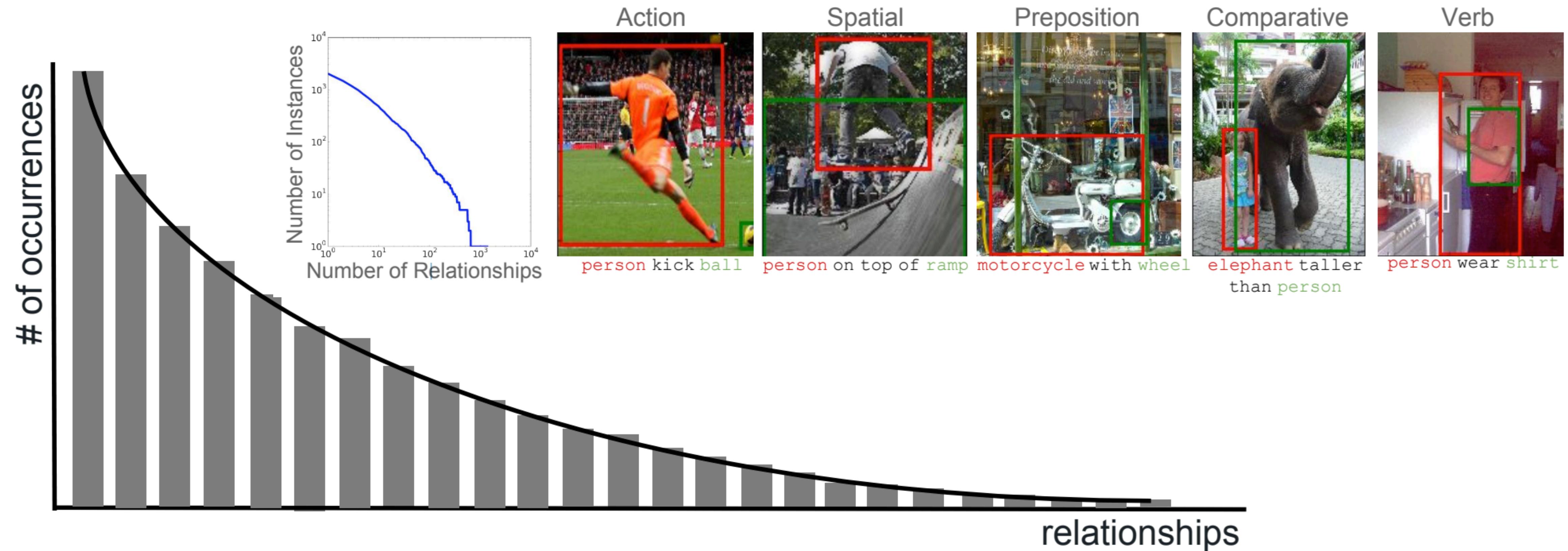
Visual Genome dataset
N = 33K
K = 42K





Challenges

- Challenge 2: Long-Tail distribution for relationships



Visual module

Proposals:



Sample: b_1 b_2



object
detector

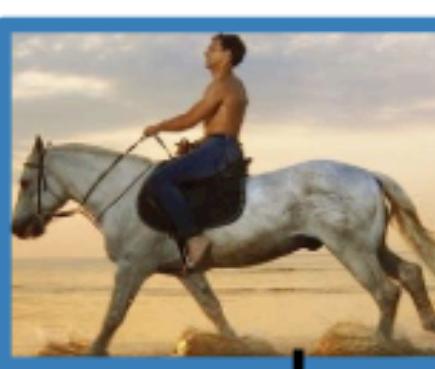
relationship
detector

$$s(o_1|b_1) \quad s(o_2|b_2)$$

$$s(r|b_1 \cup b_2)$$

$$p(T|b_1, b_2)$$

$$b_1 \cup b_2$$



$$p(T|b_1, b_2)$$

$$p(T|lang)$$

$$p(T|b_1, b_2, lang)$$

$$\underset{T}{\operatorname{argmax}}$$



Language module

o_1 : person

r : ride

o_2 : horse



$$p(T|lang)$$

$$p(T|lang)$$

Definitions:

b_1, b_2 are object proposals

$o_1, o_2 \in [\text{person}, \text{horse}, \dots]$

$r \in [\text{on}, \text{in}, \text{ride}, \text{front of}, \dots]$

T is a $\langle o_1, r, o_2 \rangle$ triple



Extensions

- Action Genome
- Home Action Genome (HOMAGE)
- EGTR (one-stage, attention graph)
- Open Vocabulary 3D Scene graph

Action Genome: Actions as Compositions of Spatio-temporal Scene Graphs, CVPR 2020
https://openaccess.thecvf.com/content_CVPR_2020/papers/

[Ji Action Genome Actions As Compositions of Spatio-Temporal Scene Graphs CVPR 2020 paper.pdf](#)

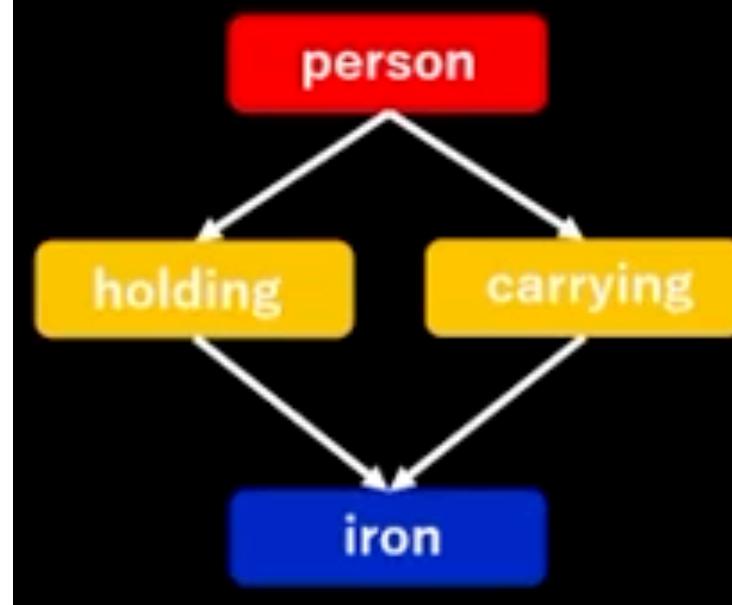
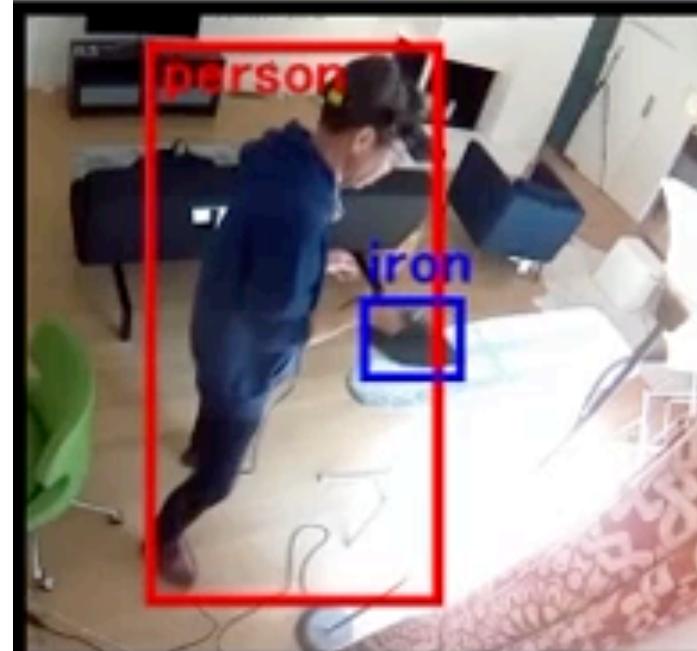
<https://homeactiongenome.org/>

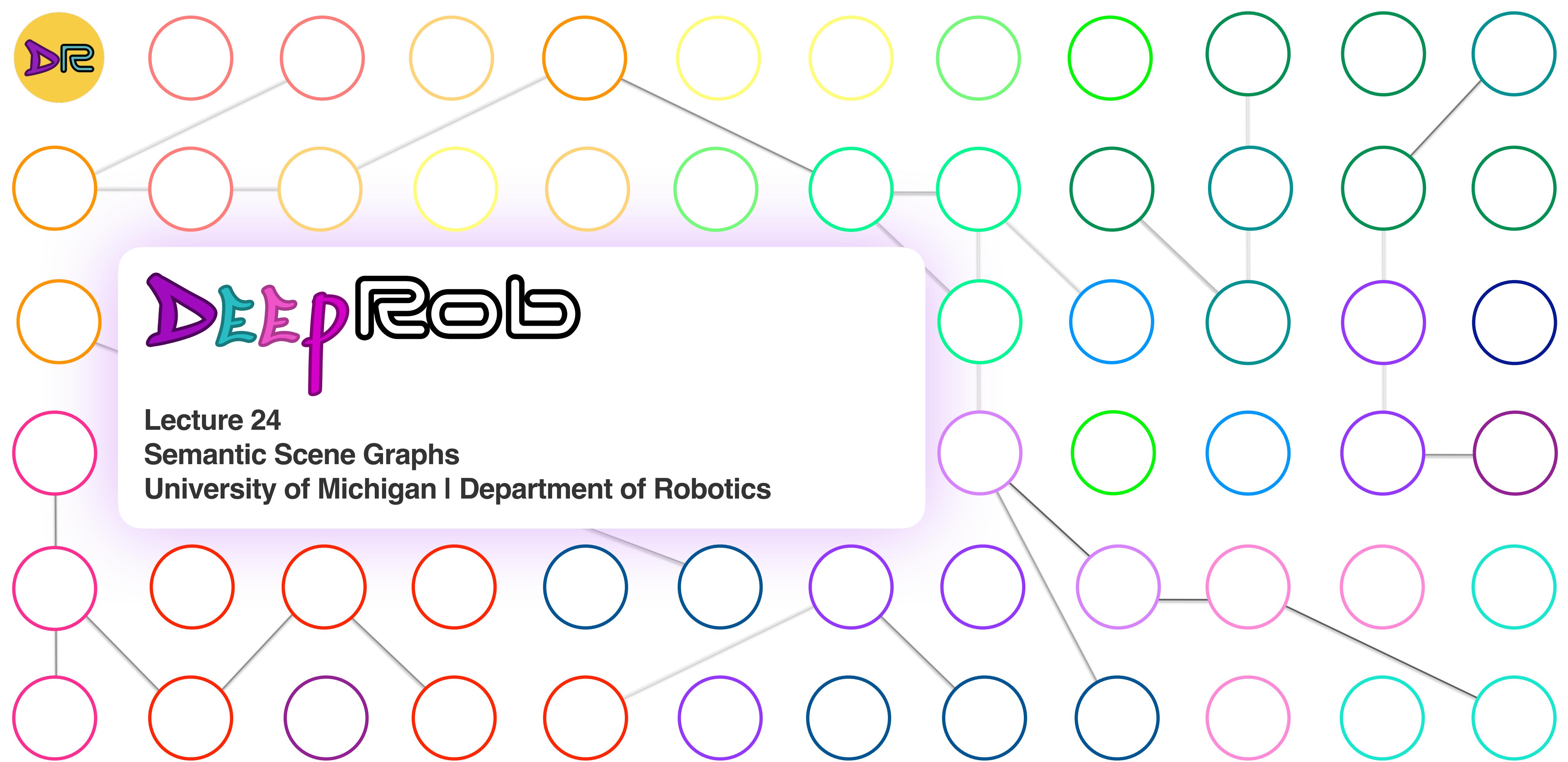
<https://arxiv.org/pdf/2404.02072.pdf> (CVPR 2024 Oral)

<https://cvpr.thecvf.com/Conferences/2024/AcceptedPapers>



Relations





DEEP ROB

Lecture 24

Semantic Scene Graphs

University of Michigan | Department of Robotics

